ITERATIVE DECODING NETWORKS WITH ITERATIVELY DATA ELIMINATING SDD AND EM BASED CHANNEL STATE ESTIMATOR

Jan Sykora[1], **Alister G. Burr**[2]

[1]Czech Technical University in Prague, FEE, Technicka 2, 166 27 Praha 6, Czech Republic, Jan.Sykora@fel.cvut.cz
[2]University of York, York YO10 5DD, U.K., alister@ohm.york.ac.uk

Abstract - The paper establishes a general framework for iterative separate CSE in general iterative decoding networks. Two particular cases of CSE are examined—SDD (Soft-Decision Directed) and EM (Expectation-Maximization) based one. Both have capabilities for exploiting the iteratively improved backward measure from the decoding network, however both exhibit different properties and provide different possibilities for iteration scenarios. An example application with simple serially concatenated code with QPSK mapping in AWGN channel with phase rotation is investigated to demonstrate differences between the algorithms in terms of MSE, ambiguity resolution, and convergence behavior.

I. INTRODUCTION

Iterative decoding algorithms received a considerable attention over the last decade. A general background can be found in [1], [2]. The core concept of the iterative signal processing is the Soft-Input Soft-Output (SISO) module [3]. Only recently, the problem of possible iterative processing was extended also for a case of the channel with unknown channel state. There are two general approaches. The first one modifies the SISO module to accommodate for unknown channel state ([4], [5]). The second one uses separate channel state estimator (CSE) utilizing the soft information measure from the decoder ([6], [7], [8]). The form of soft measure utilization ranges from purely ad-hoc approach to the more systematic ways (e.g. the expectation maximization algorithm). However in all cases, the mutual interaction between the CSE and the decoding loops has not been rigorously investigated yet.

This paper defines a rigorous general theoretical framework describing all possibilities of iterative decoding with iteratively data eliminating separate CSE. This allows to understand the mutual interactions between these two loops in a general manner and to understand a position of the separate iteratively data eliminating estimator from the perspective of optimal joint CSE and decoding. We put a particular attention on and illuminate the differences between the SDD and EM based iterative CSE.

[0]This work was supported by the Grant Agency of the Czech Republic, grant 102/02/0483/A and by the Ministry of education, youth and sports of the Czech Republic, prog. J04/98:212300014. Rev. 1.1.0 (19.5.2004).

II. SYSTEM MODEL

A. General encoding network

We assume raw information carrying data to be IID (Independent Identically Distributed) symbols d_n, $\mathbf{d} = [\ldots, d_n, \ldots]^T$. These data are fed into the first block $\mathcal{C}_1$ of a general encoding network. It is formed by arbitrary number of arbitrarily connected functional blocks—encoders modeled as Finite State Machine (FSM), interleavers, broadcasters, etc. We denote these blocks (encoders) $\mathcal{C}_s$, $s = 1, \ldots, S$. Let its input and output symbol streams be $\mathbf{c}_s = [\ldots, c_{s,n}, \ldots]^T$ and $\mathbf{q}_s = [\ldots, q_{s,n}, \ldots]^T$ respectively. The input-output relation is $\mathbf{q}_s = \mathcal{C}_s[\mathbf{c}_s]$. The output of the last encoder $\mathcal{C}_S$ is the overall output of the network and it is fed into the channel through the signal space constellation mapper $\mathbf{s} = \mathbf{s}(\mathbf{q})$. These channel symbols are denoted by $\mathbf{q} = [\ldots, q_n, \ldots]^T$.

B. Channel

The continuous-time waveform channel output is assumed to be represented by its signal space (orthonormal) expansion. The encoder is observed on the receiver side through the observation $\mathbf{x} = [\ldots, x_n, \ldots]^T$. The observation (received signal) $\mathbf{x} = \mathbf{x}(\mathbf{q}, \boldsymbol{\theta}, \mathbf{w})$ depends on channel symbols $\mathbf{q}$ and it is parametrized by channel nuisance parameters—channel state (CS) $\mathbf{w}, \boldsymbol{\theta}$.

Random channel nuisance parameters $\mathbf{w} = [\ldots, w_n, \ldots]^T$ with known a priori PDF $p(\mathbf{w})$ can be *eliminated* from $p(\mathbf{x}|\mathbf{q}, \boldsymbol{\theta}, \mathbf{w})$ to obtain $p(\mathbf{x}|\mathbf{q}, \boldsymbol{\theta})$. This eliminated conditional PDF carries all the information necessary for the construction of the detector with optimality criterion being the *average* over the eliminated parameters (e.g. the average error rate). The eliminated parameters—*eliminated channel states* (ECS) are those with close to *ergodic* behavior with respect to the channel observation period. Typically, this is additive white Gaussian noise.

The channel is said to have *independent* eliminated channel states (IECS) if $p(\mathbf{w}) = \prod_n p(w_n)$. Additionally, the channel is said to be a *memoryless channel* (MLC) if a single component of received signal depends only on a single channel symbol and single component of ECS vector $p(x_n|\mathbf{q}, \boldsymbol{\theta}, \mathbf{w}) = p(x_n|q_n, \boldsymbol{\theta}, w_n)$. The IECS-MLC assumptions directly imply $p(\mathbf{x}|\mathbf{q}, \boldsymbol{\theta}) = \prod_n p(x_n|q_n, \boldsymbol{\theta})$.

0-7803-8523-3/04/$20.00 ©2004 IEEE.

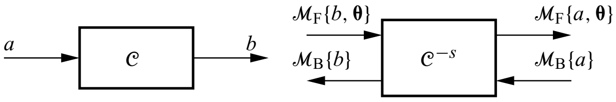

Fig. 1
PSISO module.

III. Marginalized iterative decoding network

A. Parametric Soft-Input Soft-Output module

A Parametric Soft-Input Soft-Output (PSISO) module is a soft inversion $\mathcal{C}^{-s}$ corresponding to the encoding block $b = \mathcal{C}[a]$ on general input and output quantities a and b. It is the block with two inputs and two outputs defined as follows. Forward-in, Forward-out, Backward-in, and Backward-out soft information measure is defined respectively $\mathcal{M}_{\mathrm{F}}\{b, \boldsymbol{\theta}\} = \{p(\mathbf{x}|b^{(i)}, \boldsymbol{\theta})\}_{i,\boldsymbol{\theta}}$, $\mathcal{M}_{\mathrm{F}}\{a, \boldsymbol{\theta}\} = \{p(\mathbf{x}|a^{(i)}, \boldsymbol{\theta})\}_{i,\boldsymbol{\theta}}$, $\mathcal{M}_{\mathrm{B}}\{a\} = \{p(a^{(i)})\}_i$, $\mathcal{M}_{\mathrm{B}}\{b\} = \{p(b^{(i)})\}_i$ where $\mathbf{x}$ is the observation. The forward measure carries the new information on the given quantity from the observation to the decision—hence the forward attribute. The backward measure provides a priori known information that is already available at the receiver—hence the backward attribute. Notice that there is no backward information on the uneliminated CS $\boldsymbol{\theta}$. This is implied by the independence of data and CS $\boldsymbol{\theta}$. However it holds only for *perfect* backward measure. Later we will see that this does not hold for iterative estimation of the backward measure based on non-parametrized forward measure. See Fig. 1.

B. Decoding network passing marginalized soft information measure

Marginalized PSODEM (Parametric Soft-Output Demodulator) provides $\mathcal{M}_{\mathrm{F}}\{q_n, \boldsymbol{\theta}\} = \{p(\mathbf{x}|q_n^{(i)}, \boldsymbol{\theta})\}_{n,i,\boldsymbol{\theta}}$. Individual PSISO $\mathcal{C}_s^{-s}$ in the network has forward-in soft information measure $\mathcal{M}_{\mathrm{F}}\{q_{s,n}, \boldsymbol{\theta}\} = \{p(\mathbf{x}|q_{s,n}^{(i)}, \boldsymbol{\theta})\}_{n,i,\boldsymbol{\theta}}$ and forward-out soft information measure $\mathcal{M}_{\mathrm{F}}\{c_{s,n}, \boldsymbol{\theta}\} = \{p(\mathbf{x}|c_{s,n}^{(i)}, \boldsymbol{\theta})\}_{n,i,\boldsymbol{\theta}}$. The backward-out measure $\mathcal{M}_{\mathrm{B}}\{q_{s,n}\} = \{p(q_{s,n}^{(i)})\}_{n,i}$ provided by SISO module should be understood in a sense of $q_{s,n}^{(i)}$ PDF *consistent* with the mapping (codeword structure) $p(q_{s,n}^{(i)} : \mathbf{c}_s \mapsto \mathbf{q}_s)$. In the following, we will drop this explicit notation. The SISO calculates this value from the joint PDF *respecting the codeword structure* and with subsequent marginalization for $q_{s,n}$.

Decision soft information measure on $c_{s,n}$ (and similarly for $q_{s,n}$) is $\mathcal{M}\{c_{s,n}|\boldsymbol{\theta}\} = \{p(\mathbf{x}, c_{s,n}^{(i)}|\boldsymbol{\theta})\}_{n,i,\boldsymbol{\theta}}$. Forward and backward measure is related to the decision measure by (and similarly for $\mathcal{M}\{q_{s,n}|\boldsymbol{\theta}\}$)

$$\mathcal{M}\{c_{s,n}|\boldsymbol{\theta}\} = \mathcal{M}_{\mathrm{F}}\{c_{s,n}, \boldsymbol{\theta}\}\mathcal{M}_{\mathrm{B}}\{c_{s,n}\}. \tag{1}$$

IV. Separate channel state estimator with iteratively eliminated data

A. Data elimination in decoding network

As a alternative to the *joint* data detection and CSE problem we can build a decoding network with separately working CSE and data detector. The Maximum Likelihood (ML) CSE is derived from the soft information measure with eliminated random data. Assume the joint decision measure $\mathcal{M}\{\mathbf{d}|\boldsymbol{\theta}\} = \{p(\mathbf{x}, \mathbf{d}^{(i)}|\boldsymbol{\theta})\}_{i,\boldsymbol{\theta}}$. Data elimination leads to decision measure on CS with eliminated data

$$\mathcal{M}\{|\boldsymbol{\theta}\} = \left\{\sum_{\mathbf{d}} p(\mathbf{x}, \mathbf{d}|\boldsymbol{\theta})\right\}_{\boldsymbol{\theta}} = \{p(\mathbf{x}|\boldsymbol{\theta})\}_{\boldsymbol{\theta}}\,. \tag{2}$$

Data $\mathbf{d}$ are the raw information data usually being IID $p(\mathbf{d}) = \prod_n p(d_n)$ with known (usually uniform) a priori PDF.

A similar relationship holds for an arbitrary intermediate encoding stage with codewords $\mathbf{c}$

$$\mathcal{M}\{|\boldsymbol{\theta}\} = \left\{\sum_{\mathbf{c}:\mathbf{d}\mapsto\mathbf{c}} p(\mathbf{x}|\mathbf{c}, \boldsymbol{\theta})p(\mathbf{c})\right\}_{\boldsymbol{\theta}} \tag{3}$$

where $\mathbf{c} : \mathbf{d} \mapsto \mathbf{c}$ means all $\mathbf{c}$ consistent with the mapping $\mathbf{d} \mapsto \mathbf{c}$, i.e. the structure of valid codewords $\mathbf{c}$. Also the *joint* PDF $p(\mathbf{c})$ captures this implicitly. This means that the choice of the encoding stage (codewords) over which we perform the elimination is *arbitrary*, provided that we eliminate over valid codewords, i.e. we respect the *structure* of the code.

The data decisions with separate CS estimator with data elimination are then

$$\hat{\mathbf{d}} = \arg\max_{\check{\mathbf{d}}} \mathcal{M}\{\check{\mathbf{d}}|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}\} \tag{4}$$

where ML CSE is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_{\mathbf{d}} \mathcal{M}\{\mathbf{d}|\check{\boldsymbol{\theta}}\} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_{\mathbf{c}:\mathbf{d}\mapsto\mathbf{c}} p(\mathbf{x}|\mathbf{c}, \check{\boldsymbol{\theta}})p(\mathbf{c}). \tag{5}$$

B. Factorization of the elimination with perfect backward measure

The *joint* PDF on the channel symbols (codewords) $\mathbf{q}$ (or any intermediate code $\mathbf{c}$) is *not directly available* in the iterative *marginalized* decoding network. It produces the backward soft information *marginalized* measure estimate on channel symbol q_n at each iteration.

Assume for a moment that the *perfect* backward marginalized measure $\mathcal{M}_{\mathrm{B}}\{q_n\}$ is available. In order to apply the elimination (3) we must (a) reconstruct the joint PDF to respect the code structure $\mathbf{d} \mapsto \mathbf{q}$ *or* (b) use the marginalized PDF in such a way that is equivalent to the use of the joint PDF. The equivalent use means such that naturally leads to marginalization even when correctly joint PDF is assumed at the beginning. That is, if we succeed to manipulate the elimination equation (3) in such a way that it would require only marginalized PDF. This would be a *factorized* form of the elimination.

C. Equivalence of iterative CSE with direct elimination

Assuming that the iterative CSE in the iterative decoding network converges and we succeeded to find its factorized form then it would *not* provide any performance advantage over the direct elimination at the raw data $\mathbf{d}$ level (of course both respecting the correct mapping $\mathbf{d} \mapsto \mathbf{q}$).

D. Multidimensional feed-back system—State variables and iteration index vector

The iterative decoding marginalized soft information measure passing network is formed from PSISO modules that use on the backward-in node current estimation (state) of the information measure provided by the previous firing of the neighboring module. This information measure must be stored as the feed-back network *state variable*. There will be $S-1$ memory blocks (a priori information on raw data $\mathbf{d}$ is constant during the iterative process), i.e. the feed-back systems will be $S-1$ dimensional. The iteration step of this discrete feed-back system can be described by $S-1$ dimensional index $\mathbf{m} = [m_1, \ldots, m_{S-1}]^T$. A particular iteration index m_s is increased by each update of the corresponding s-th memory module at the backward-in input of the $\mathcal{C}_s^{-s}$ module. In the case of separate CSE, the decoding network bases its soft information measures on the current k-th CSE iteration. Thus, we must use overall iteration index $(k, \mathbf{m})$ where k denotes iterations over the CSE and $\mathbf{m}$ denotes the iteration within the decoding network.

E. Non-parametric soft information measure in iterative network with separate CSE

1) Forward soft information measure: The soft output demodulator (SODEM) of the network with separate CSE provides *forward* soft information measure *only on channel symbols*—the measure is *not parametrized*, i.e. it does not convey the information on CS for its final decision later on. However, it was obtained by assuming CS estimation $\hat{\boldsymbol{\theta}}^{k,\mathbf{m}}$ at current iteration step $k, \mathbf{m}$ in the iteration. The soft measure at the SODEM output can be interpreted as a *cut* in the 2-dimensional joint measure $\mathcal{M}_{\mathrm{F}}\{q_n, \boldsymbol{\theta}\}$ over the hyperplane $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{k,\mathbf{m}}$ based on an estimate corresponding to the $k, \mathbf{m}$ iteration

$$\hat{\mathcal{M}}_{\mathrm{F}}^{k,\mathbf{m}}\{q_n\} = \{p(\mathbf{x}|q_n^{(i)}, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}})\}_{n,i} = \{\hat{p}^{k,\mathbf{m}}(\mathbf{x}|q_n^{(i)})\}_{n,i}. \quad (6)$$

The same holds for all forward soft information measures in the network. They are estimates of the measure only on corresponding symbols, not on the CS. The estimates are understood in a sense of the cut of full joint measure at the estimate hyperplane $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{k,\mathbf{m}}$. The forward measure changes with all iterations over any of the iteration indices $k, \mathbf{m}$.

2) Backward soft information measure: Unlike for the perfect backward measure case (1), the backward measure *iterative estimate* depends on current CSE $\hat{\boldsymbol{\theta}}^{k,\mathbf{m}}$

$$\hat{\mathcal{M}}_{\mathrm{B}}^{k,\mathbf{m}}\{c_n\} = \hat{\mathcal{M}}^{k,\mathbf{m}}\{c_n\}/\hat{\mathcal{M}}_{\mathrm{F}}^{k,\mathbf{m}}\{c_n\}. \quad (7)$$

This CSE determines the hyperplane cut of the PSODEM forward measure output and therefore it affects current iteration run of the detection network. The choice of the cut $\hat{\boldsymbol{\theta}}^{k,\mathbf{m}}$ should not be confused with the actual CS value $\boldsymbol{\theta}$ which is independent with data and therefore is not affecting the backward measure. If $\hat{\boldsymbol{\theta}}^{k,\mathbf{m}} = \boldsymbol{\theta}$ was true (perfect estimate), than the backward measure would be independent with the true CS value $\boldsymbol{\theta}$.

V. Soft Decision Directed and Expectation-Maximization based CSE

A. Soft Decision Directed approach

A soft decision directed (SDD) approach (for an example application see [8]) is an *iterative* implementation of the data elimination principle (3). The encoding stage, at which the elimination is performed, is usually directly the channel symbols $\mathbf{q}$ level.

1) Approximate factorization of the elimination: At the output to the CS estimator module, the marginalized decoding network does not provide the joint PDF $p(\mathbf{q})$. Instead, it provides only marginalized iterative *approximations* of $\hat{\mathcal{M}}_{\mathrm{B}}^{k,\mathbf{m}}\{q_n\} = \{\hat{p}^{k,\mathbf{m}}(q_n^{(i)})\}_{n,i}$ at the $(k, \mathbf{m})$-th iteration. The factorized form of the expectation (5) using the *estimates* of backward measure must rely on the *approximation* (with explicit and implicit notation—see Sec. III-B)

$$\hat{p}^{k,\mathbf{m}}(\mathbf{q}) \approx \prod_n \hat{p}^{k,\mathbf{m}}(q_n : \mathbf{d} \mapsto \mathbf{q}) = \prod_n \hat{p}^{k,\mathbf{m}}(q_n). \quad (8)$$

Moreover, it is additionally affected by the fact that the backward measure is *not perfect* but only approximate.

2) Final iterative solution: Assuming a *IECS-MLC* channel from the perspective of channel symbols $\mathbf{q}$ and using the approximation above, we arrive to the final iterator

$$\hat{\boldsymbol{\theta}}^{k+1,\mathbf{m}} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_n \ln\left(\sum_{q_n} p(x_n|q_n, \check{\boldsymbol{\theta}})\hat{p}^{k,\mathbf{m}}(q_n)\right). \quad (9)$$

We also applied logarithm to get the LLF. Notice that the right-hand side of (9) *depends* on the iteration index k, i.e. iterations over the CS estimate, only through $\hat{p}^{k,\mathbf{m}}(q_n)$ which is fixed for given fixed $\mathbf{m}$. It therefore only make sense to iterate *simultaneously* at one step both k and $\mathbf{m}$ index loops. The iteration over k index loop does not change the CS estimate. An example network with iterative elimination and iterative decoding is on Fig. 2.

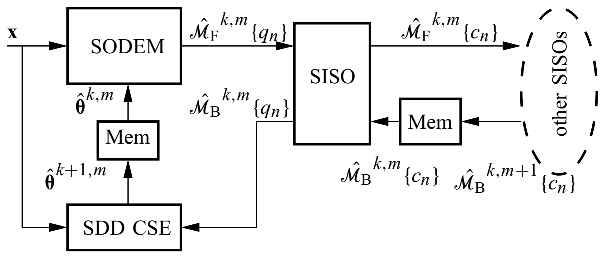

Fig. 2
Iterative non-parametric SDD decoding network—multidimensional feed-back system. An example of serial concatenated network.

B. Expectation-Maximization approach

1) Basic principle: An expectation-maximization (EM) procedure (see [9], [10]) is an iterative method for obtaining maximum likelihood (ML) estimation. It is based on the following relatively simple idea. Assume that $\boldsymbol{\theta}$ is unknown deterministic parameter that we want to build the ML estimator for. Assume that $\mathbf{x}$ is available observation. Also assume that there is observation $\mathbf{y}$ that is *not* available to the estimator. An union of these two $\mathbf{x} \cup \mathbf{y}$ is called a complete observation. Assume that the estimator based on the log-likelihood function (LLF) $\ln p(\mathbf{x}|\boldsymbol{\theta})$ is feasible however another one based on LLF $\ln p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta})$ is *easier* to build. But the observation $\mathbf{x}, \mathbf{y}$ is not available hence we replace it by an *approximation* marginalizing the unavailable observation $\mathbf{y}$

$$\ln p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) \approx \mathrm{E}_{\mathbf{y}|\mathbf{x},\hat{\boldsymbol{\theta}}^k}\left[\ln p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta})\right]. \tag{10}$$

The marginalization (expectation) can be interpreted as obtaining the average expected value of LLF $\ln p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta})$ provided that we know the best currently available information about quantities bound together by this LLF. It is the available observation $\mathbf{x}$ and the current guess of the parameter $\hat{\boldsymbol{\theta}}^k$. We are replacing $\ln p(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta})$ by its expected value over unavailable observation extrapolated from the information on available observation and current guess of the parameter. The EM iteration is then described by

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\check{\boldsymbol{\theta}}} \mathrm{E}_{\mathbf{y}|\mathbf{x},\hat{\boldsymbol{\theta}}^k}\left[\ln p(\mathbf{x}, \mathbf{y}|\check{\boldsymbol{\theta}})\right]. \tag{11}$$

2) Application to CS estimation: An application of the EM algorithm to the CS estimation problem in data communication is usually done (see [6]) by setting the unavailable observation equal to the data $\mathbf{y} = \mathbf{d}$. The estimator based on observation $\mathbf{x}, \mathbf{d}$ would be definitively easier to build—it is in fact the data aided (DA) estimator. The EM approach marginalizes the knowledge of true transmitted data by iterations

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\check{\boldsymbol{\theta}}} \mathrm{E}_{\mathbf{d}|\mathbf{x},\hat{\boldsymbol{\theta}}^k}\left[\ln p(\mathbf{x}, \mathbf{d}|\check{\boldsymbol{\theta}})\right]. \tag{12}$$

Realizing that the data are independent with the CS we get

$$p(\mathbf{x}, \mathbf{d}|\boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{d}, \boldsymbol{\theta})p(\mathbf{d}|\boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{d}, \boldsymbol{\theta})p(\mathbf{d}). \tag{13}$$

A substitution of this result into (12) gives

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_{\mathbf{d}} \ln p(\mathbf{x}|\mathbf{d}, \check{\boldsymbol{\theta}})p(\mathbf{d}|\mathbf{x}, \hat{\boldsymbol{\theta}}^k). \tag{14}$$

The averaging can be, similarly as in the data elimination case, done at arbitrary encoding stage (e.g. at the level of channel symbols $\mathbf{q}$)

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_{\mathbf{q}:\mathbf{d}\mapsto\mathbf{q}} \ln p(\mathbf{x}|\mathbf{q}, \check{\boldsymbol{\theta}})p(\mathbf{q}|\mathbf{x}, \hat{\boldsymbol{\theta}}^k). \tag{15}$$

Compare the expectation in this equation with the expectation operation in (3). The averaging in EM algorithm uses the a posteriori PDF.

3) Factorization of the expectation with perfect channel symbols measure: The expectation in (15) over the joint PDF must be equivalently transformed (if possible) in the form using only *marginalized* PDF which is the only available quantity in marginalized iteration network.

Similarly as for the SDD case, we must assume a *IECS-MLC* channel from the perspective of channel symbols $\mathbf{q}$: $p(\mathbf{x}|\mathbf{q}, \boldsymbol{\theta}) = \prod_n p(x_n|q_n, \boldsymbol{\theta})$. Then the expectation in (15) is

$$\begin{aligned}&\sum_{\mathbf{q}:\mathbf{d}\mapsto\mathbf{q}} \ln p(\mathbf{x}|\mathbf{q}, \check{\boldsymbol{\theta}})p(\mathbf{q}|\mathbf{x}, \hat{\boldsymbol{\theta}}^k)\\ &= \sum_n \sum_{q_n} \ln p(x_n|q_n, \boldsymbol{\theta}) \sum_{\mathbf{q}:q_n,\mathbf{d}\mapsto\mathbf{q}} p(\mathbf{q}|\mathbf{x}, \hat{\boldsymbol{\theta}}^k).\end{aligned} \tag{16}$$

In the previous equation, the marginalizing expression $p(q_n|\mathbf{x}, \hat{\boldsymbol{\theta}}^k) = \sum_{\mathbf{q}:q_n,\mathbf{d}\mapsto\mathbf{q}} p(\mathbf{q}|\mathbf{x}, \hat{\boldsymbol{\theta}}^k)$ is due to a condition $\mathbf{q} : q_n, \mathbf{d} \mapsto \mathbf{q}$ *correctly respecting the code structure*. This marginalization is exactly what the SISO module provides.

The estimator is thus

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_n \sum_{q_n} \ln p(x_n|q_n, \boldsymbol{\theta})p(q_n|\mathbf{x}, \hat{\boldsymbol{\theta}}^k). \tag{17}$$

The a posteriori PDF can be easily get from the hyperplane $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^k$ cut of the forward joint measure $\mathcal{M}_{\mathrm{F}}\{q_n, \boldsymbol{\theta}\}$ and the backward measure on $\mathcal{M}_{\mathrm{B}}\{q_n\}$

$$p(q_n|\mathbf{x}, \hat{\boldsymbol{\theta}}^k) = \frac{p(\mathbf{x}|q_n, \hat{\boldsymbol{\theta}}^k)p(q_n)}{\sum_{q_n} p(\mathbf{x}|q_n, \hat{\boldsymbol{\theta}}^k)p(q_n)}. \tag{18}$$

4) A posteriori symbol PDF with iterative channel symbols backward measure: In the iterative decoding network, only iterative estimates of the backward measure $\hat{\mathcal{M}}_{\mathrm{B}}^{k,\mathbf{m}}\{q_n\}$ are available. This measure replaces the true a priori PDF in the final iterative synchronizer. The a posteriori symbol PDF estimate is thus

$$\hat{p}^{k,\mathbf{m}}(q_n|\mathbf{x}, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}}) = \frac{p(\mathbf{x}|q_n, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}})\hat{p}^{k,\mathbf{m}}(q_n)}{\sum_{q_n} p(\mathbf{x}|q_n, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}})p^{k,\mathbf{m}}(q_n)}. \tag{19}$$

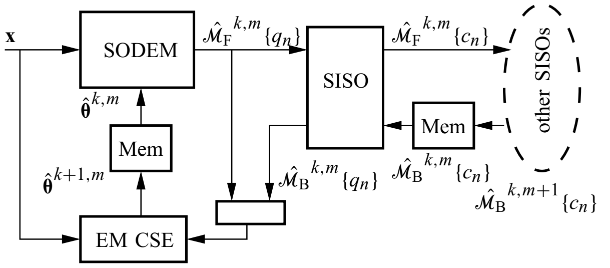

Fig. 3
Iterative non-parametric EM decoding network—multidimensional feed-back system. An example of serial concatenated network.

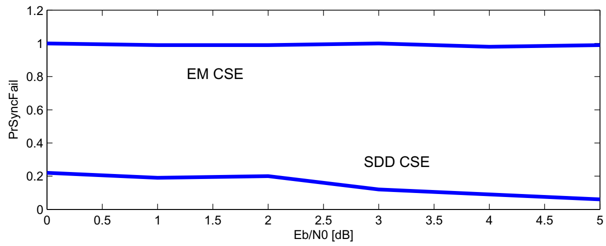

Fig. 4
Probability of synchronization failure (ambiguity domain resolution). True channel phase is 80 degrees.

5) Final iterative solution: The final CSE iterator is

$$\hat{\boldsymbol{\theta}}^{k+1,\mathbf{m}} = \arg\max_{\check{\boldsymbol{\theta}}} \sum_n \sum_{q_n} \ln p(x_n|q_n, \check{\boldsymbol{\theta}}) \hat{p}^{k,\mathbf{m}}(q_n|\mathbf{x}, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}}) \tag{20}$$

where the symbol a posteriori PDF estimation is given by (19). Compare this result with SDD case (9). Unlike the SDD case, the right-hand side depends on k even for fixed $\mathbf{m}$ and the iteration loops can run *independently*. The possibility of iterating over k and improving the estimation can save the number of necessary runs of computationally expensive decoding iteration (Forward-Backward Algorithm). An example network with iterative EM CSE is on Fig. 3.

VI. Example application—simulation results

A. Demonstration system description

As a demonstration system, we chose a simple serially concatenated encoding network. The outer code $\mathbf{d} \mapsto \mathbf{c}$ is defined by the output $c_n(d^{(i)}, \sigma^{(k)}) = c^{(\mathbf{C}_{i,k})}$ and the state equation $\sigma_{n+1}(d^{(i)}, \sigma^{(k)}) = \sigma^{(\boldsymbol{\Sigma}_{i,k})}$ where $\mathbf{C} = [1, 3; 2, 4]$, $\boldsymbol{\Sigma} = [1, 1; 2, 2]$ (Matlab-like row-wise notation). The inner code $\tilde{\mathbf{c}} \mapsto \mathbf{q}$ is $q_n(\tilde{c}^{(i)}, \sigma^{(k)}) = q^{(\mathbf{Q}_{i,k})}$, $\sigma_{n+1}(d^{(i)}, \sigma^{(k)}) = \sigma^{(\boldsymbol{\Sigma}_{i,k})}$ where $\mathbf{Q} = [1, 2, 3, 4; 1, 2, 3, 4; 1, 2, 3, 4; 1, 2, 3, 4]$, $\boldsymbol{\Sigma} = [1, 2, 3, 4; 2, 3, 4, 1; 3, 4, 1, 2; 4, 1, 2, 3]$. The codes are separated by the UMTS340 interleaver $\tilde{\mathbf{c}} = \boldsymbol{\Pi}(\mathbf{c})$ directly applied on 4-ary symbols. The signal space QPSK constellation mapping is $s_n(q^{(1)}) = 1$, $s_n(q^{(2)}) = j$, $s_n(q^{(3)}) = -j$, $s_n(q^{(4)}) = -1$. The signal is passed through the AWGN channel with phase rotation φ. The noise has the complex envelope power spectral density $2N_0$. The signal space expansion model of the received signal is $x_n = s_n e^{j\varphi} + w_n$. The received bit energy to N_0 ratio is γ_B.

B. Performance of the iterative SDD and EM CSE

We have simulated the performance of the SDD and EM based phase estimator. A MSE (Mean Square Error) convergence behavior and the capability of the phase ambiguity resolution were investigated. The phase ambiguity is is given by the rotational invariance of the QPSK constellation. If there was no coding the invariance would be $2\pi/4$. However a synchronizer properly utilizing the codeword structure can reduce this ambiguity if the valid codeword space has lower rotational ambiguity. All numerical results for the EM case assume 5 CSE iterations per one decoding iteration and the initial estimate $\hat{\varphi}^{0,0} = 0$.

1) Iteration scenarios: The combined iterative decoding and synchronization loop can iterate over k and m indices. In the case of SDD CSE only joint iteration is possible. However for the EM CSE there is possible to iterate several times over the synchronizer loop (index k) per one decoding iteration (index m). This of course saves computationally expensive Expectation-Maximization Algorithm runs.

2) Ambiguity resolution—synchronization failure: The EM CSE demonstrates very strong dependence on the choice of the initial guess. If the initial guess does *not* lie in the *correct ambiguity domain*, the algorithm fails to correct this even in higher iterations and remains in the wrong domain with probability close to 1. See Fig. 4.

This behavior is caused by averaging the *logarithm* of the PDF, i.e. the *distances* between the received signal and rotated points of constellation, in the elimination (20). The identity $\sum_{i=0}^{M-1} ||x|e^{j\varphi} - e^{j\check{\varphi}} e^{j\frac{2\pi}{M}i}|^2 = M(|x| + 1)$ holds for a PSK type constellation. This means that for an initial iteration with uniform backward measure, the estimator objective function (20) dependence on the phase is caused only by the term $p(\mathbf{x}|q_n, \hat{\boldsymbol{\theta}}^{k,\mathbf{m}})$ which strongly depends on the initial guess amplifying the objective function in a wrong domain. Compare the behavior of SDD and EM CSE on Fig. 5. Especially notice the different behavior for the synchronization failure case. The SDD algorithm has much better chance to get into the correct ambiguity domain unlike the EM algorithm which stays in the domain of initial estimate with probability close to 1.

3) Mean Square Error convergence: The EM CSE has smoother MSE convergence than SDD CSE provided that its initial estimate was in the correct ambiguity domain. See Fig. 5.

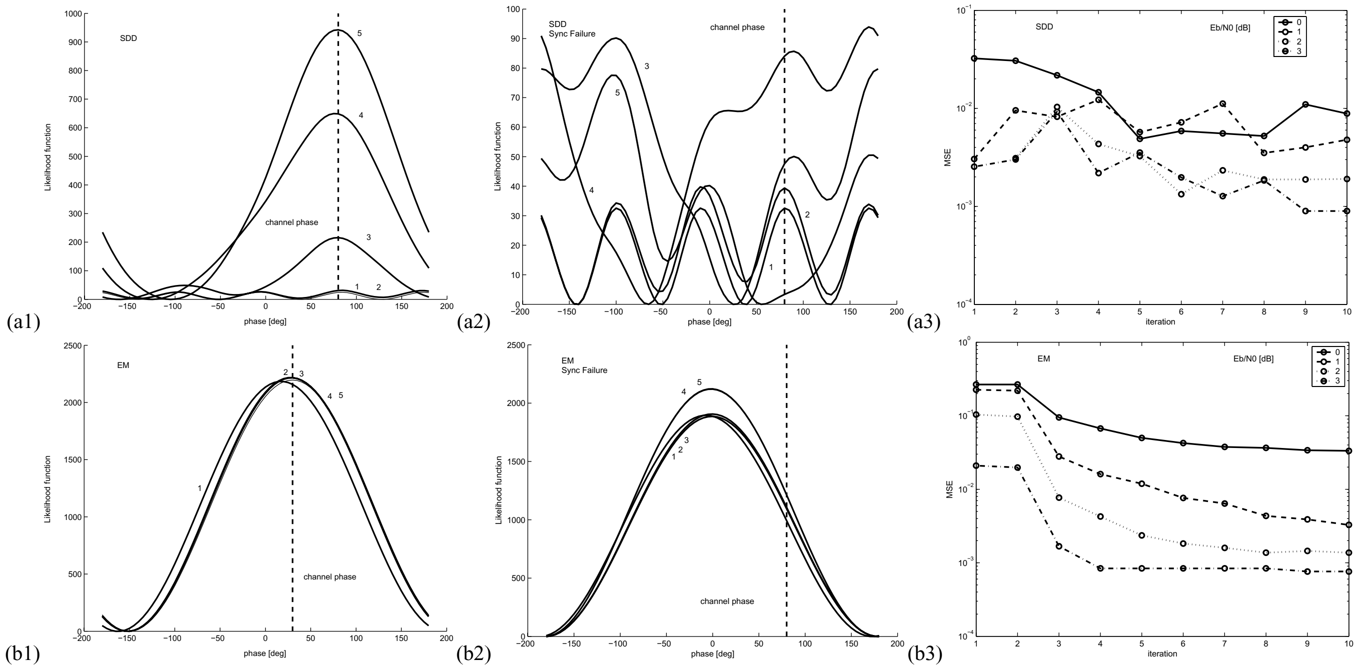

Fig. 5

Numerical results for (a) SDD and (b) EM CSE. The estimator objective function for different iterations and (1) successful ambiguity resolution, (2) synchronization *failure*. True channel phase is shown as a dashed line. The decoding loop iteration number is a parameter, $\gamma_B = 2\,[\mathrm{dB}]$. The MSE (3) as a function of the iteration number.

VII. Conclusions

The general framework for iterative CSE was established. We investigated MSE and the estimator objective function behavior as a function of iteration number. The most notable is a strong dependence of EM CSE on the initial estimate (unlike for the SDD case) resulting in high probability of synchronization failure if the ambiguity is not resolved a priori. On the other side, the EM CSE convergence is smoother than in the SDD case. Important difference between SDD and EM is also in the different possible iteration scenarios. In the EM case, the iterations over the CSE with given fixed iteration step of the decoding network improve the estimate quality. It allows to save a computationally expensive runs of expectation-maximization algorithm.

References

[1] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding: Turbo-codes,” *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.

[2] K. Chugg, A. Anastasopoulos, and X. Chen, *Iterative Detection: Adaptivity, Complexity Reduction, and Applications*. Kluwer Academic Publishers, 2001.

[3] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, “A soft-input soft-output APP module for iterative decoding of concatenated codes,” *IEEE Commun. Lett.*, vol. 1, pp. 22–24, Jan. 1997.

[4] A. Anastasopoulos and K. M. Chugg, “Adaptive iterative detection for phase tracking in turbo-coded systems,” *IEEE Trans. Commun.*, vol. 49, pp. 2135–2144, Dec. 2001.

[5] A. Anastasopoulos and K. M. Chugg, “Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty,” *IEEE Trans. Commun.*, vol. 48, pp. 1638–1649, Oct. 2000.

[6] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, “Turbo synchronization: an EM algorithm interpretation,” in *Proc. IEEE Internat. Conf. on Commun. (ICC)*, 2003.

[7] Z. Baranski, A. M. Haimovich, and J. Garcia-Frias, “EM-based iterative receiver for space-time coded modulation with noise variance estimation,” in *Proc. IEEE Global Telecommun. Conf. (GlobeCom)*, 2002.

[8] L. Zhang and A. Burr, “Phase estimation with the aid of soft output from Turbo decoding,” in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Fall 2001.

[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, Dec. 1977.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.